CHAPTER 2

LEXICAL DEVELOPMENT IN LANGUAGE ACQUISITION AND LEARNING

M. Krzysztof Szymczak

2.1 Introduction

This chapter reviews a number of key issues related to the acquisition of L2 vocabulary. It presents three distinct senses of **lexicon** and different meanings of **word**, which might be thought of as a **token**, a word **type**, or a **lexeme**. It emphasizes the importance of lexical units consisting of more than one word and touches on issues related to the size of lexicon. Several aspects of what it means to know a word are discussed: spoken and written form, word parts, connecting form and meaning, concepts and referents, associations, grammatical functions, collocations, and constraints on use. The difference between productive and receptive word knowledge is recognized, and important aspects of testing word knowledge are pointed out. The notion of **mental lexicon** is elaborated on and related to such models of lexical processing as the **logogen model**, the **cohort model**, the **lexical search model**, and Levelt's **blueprint for the speaker**, followed by an approach to mental processing called **connectionism** and **spreading/ interactive activation** models. Given that one major question about the nature of the bilingual lexicon concerns the degree of integration and separation of L1 and L2 lexicons, we shall explore some essential aspects of lexical development in both L1 and L2, the latter against the backdrop of the prior discusssion of what it means to know a word.

2.2 Three meanings of lexicon

One often thinks of language, the way we hear it spoken or read its written form, as consisting of basic building blocks called words, which are combined into utterances by means of grammar rules in order to convey some kind of meaning and perform a communicative function. It is possible to communicate quite a lot knowing only vocabulary, but it is not possible to communicate using grammar alone. After all, going abroad, we are much less likely to take a grammar book of an unfamiliar language than a dictionary or a phrase book, which are primarily sources of words or phrases.

Moving away from Chomsky's (1957) view of the supremacy of syntax (grammar), expressed in *Syntactic Structures,* researchers have come a long way towards emphasizing the importance of lexis (vocabulary). A key principle underlying Lewis' *Lexical Approach* is that "[l]anguage consists of grammaticalised lexis, not lexicalized grammar" (Lewis 1993: vi).

Central to any discussion on lexical development is the notion of **lexicon**. *Longman Dictionary of Contemporary English* (2006) defines it as follows:

> **1 the lexicon** *technical* all the words and phrases used in a language or that a particular person knows
>
> **2** [countable] an alphabetical list of words with their meanings, especially on a particular subject or in a particular language: *a lexicon of geographical terms*.

For our purposes, let us emphasize three distinct uses of the term, corresponding to three key notions: a language, a language user, and a book. Thus, a lexicon will denote the vocabulary of a given language, the vocabulary known by a specific person, or a book listing words in alphabetical order, a dictionary. To refer specifically to a language user's knowledge of vocabulary, the term **mental lexicon** is used.

We take it for granted that dictionaries are arranged alphabetically. However, some organize vocabulary according to meaning, e.g. *Roget's Thesaurus* in its classic form (not just an alphabetical dictionary of synonyms) or *Longman Lexicon of Contemporary English* (McArthur 1981). Vocabulary in these books is divided into major categories, which are further subdivided into more and more specific sections. If we want to find *orange* in the *Longman Lexicon*, we first need to decide what meaning we want – "fruit" or "colour". If we want the first, then using the top-down method we have to start with *Life and living things,* then find *Plants generally*, go to *Kinds of fruit* and *Citrus fruit*, to finally find *orange* side by side with *tangerine*, *lemon*, and *grapefruit*. Following a semantic track in this kind of search is time-consuming, and not always very successful.

To facilitate the search, the *Longman Lexicon* has an alphabetical index, where the word *orange* is tagged with a symbol (A150), which serves as a form of address, at which to find the word in the first part of the book. To accomplish our two-stage search procedure, we need to use the orthographic form of a word in order to find the word in its semantic context. On balance, finding a word in a dictionary is easy if we know its spelling, but it is more difficult if we start with meaning. Would it be possible to look up a word if we knew only its pronunciation? Not unless we knew its sound-spelling relationship. As yet, there are no marketed dictionaries listing words according to their phonological form.

2.3 What is a word? Tokens, types, and lexemes (or lemmas?)

One possible way of defining a word is to describe it as the smallest semantic unit (a unit carrying meaning) "that can move around in an utterance" (Clark 1993: 2). However, the exact meaning of "the term *word* will depend very much on the level of abstraction at which a given speaker/writer is operating" (Singleton 1999: 10). Singleton illustrates the point with the example *Going, going, gone* (a phrase used while accepting the final bid at an auction, or when describing a home run in baseball). How many words does the phrase consist of? The number differs, depending on whether we think of words in terms of **tokens**, **types**, or **lexemes**. The phrase contains three words understood as tokens or running words (*going, going, gone*), two words understood as word types (*going, gone*), or just one word, one lexeme (the verb *go*). A lexeme is an "abstract unit based on a collection of forms thus seen as constituting in some sense a single lexical entity" (Singleton 1999: 10). A lexeme comprises different word forms, which are usually the same part of speech; thus, *go, went, gone, goes, going* are forms of the same lexeme. A lexicon is made up of lexemes. (Confusing as this might be at this point, some researchers, e.g. Nation (2001), use lemma for what we have just defined as a lexeme. The term lemma is used differently in lexicography and psycholinguistics.)

By convention, one form of a lexeme is used as a headword in a dictionary to represent the whole lexeme, and that form is called the citation form. *Go* is the citation form of the lexeme discussed above. In Polish, a noun lexeme is represented by a singular nominative form (e.g. *stół*), a verb lexeme is by an infinitive form (e.g. *pisać*), and an adjective lexeme is cited by means of a singular nominative masculine form (e.g. *sosnowy*). Citation forms are printed alphabetically in a dictionary as headwords, which helps us locate a particular entry.

Apart from the level of abstraction, what we mean by a word will also depend on "the linguistic 'level(s)' being discussed, and the extent to which semantic content is being treated as criterial" (Singleton 1999: 10). Applying orthographic

criteria, whereby a word is defined as a series of letters preceded and followed by a blank space (or a punctuation mark), we will describe the word *bushes* as a sequence of six letters: *b*, *u*, *s*, *h*, *e*, and *s*. In phonological terms, the word consists of five phonemes: /b/, /ʊ/, /ʃ/, /ɪ/ and /z/. Morphosyntactically, it consists of the noun stem *bush* and the plural suffix *-es*, and at a semantic level, it is similar to *shrubs* and refers to plants which grow up from the ground and have a lot of thin branches. The definition of word at the beginning of this section is a semantico-grammatical one.

With regard to the semantic content of words, we distinguish between content words (or lexical words) and function words (or grammatical words). Content words (nouns, verbs, adjectives, and adverbs) carry essential meaning, regardless of context (*walk*, *orange*, *quickly*), while function words (e.g. articles, prepositions, and pronouns) express grammatical relationships (*a*, *at*, *them*).

Morphology studies internal word structure and word formation, using the notion of a morpheme. Morphemes include stems and affixes, which are added to stems. An affix is called a *prefix* when it is added to the beginning of a stem, or a suffix when it comes at the end. The word *unspeakable* consists of three morphemes: the stem *speak*, the prefix *un-*, and the suffix *-able*. We distinguish between inflected and derived forms of words. Inflection produces a different form of the same word, which is still the same part of speech, e.g. *snob* → *snobs*, while derivation changes the meaning (and often the part of speech) and produces a new word, e.g. *snob* → *snobbery/snobbish*. Inflections in English comprise plural and possessive for nouns; third person singular present tense, past tense, past participle, and *-ing* for verbs; and comparative and superlative for adjectives. In Polish, an inflecting language, there is an enormous number of inflected word forms traditionally associated with declension (nouns, adjectives, and pronouns) and conjugation (verbs). Inflected forms have predictable meaning; however, derivation may produce dramatic changes in meaning (e.g. the verb *tell* means "communicate", but the derived noun *teller* means "bank worker"). Derivation is an extremely productive process of word-formation and a major part of dictionary entries contains derived forms of words.

Since there is no universal definition of word, lexicographers have to make certain decisions, and those decisions are not always alike, which influences the way dictionary entries are constructed and how words or lexical units are counted. Should acronyms, product names (*Mercedes*, *Clorox*), people's names, and geographical names be counted as words and included in a dictionary? What about vocalizations like *uh huh* ("yes"), *uh uh* ("no"), or *uh oh* ("gosh")? The word *walker* can mean different things: (1) someone who walks for exercise, (2) a metal frame helping sick or old people to walk (American use, corresponding to the British *Zimmer frame*), (3) a frame helping a baby to move around before it can walk, also called a *baby walker*. Should *walker* be

treated as one word, two words ("person" or "frame"), or three? Relying only on the spoken form, would we decide that *nonetheless* should be treated as one word and *none the worse* as three? In terms of meaning and syntactic function, they both constitute units.

2.4 Multi-word items

There is strong evidence against viewing lexis exclusively in terms of single word units, resulting from the fact that native speakers acquire and use multi-word chunks, which function as single, unanalysed items or word-like units. Stored and retrieved from memory as whole items, they considerably increase fluency (Pawley and Syder 1983: 192), unlike those sentence parts which have to be constructed from several elements from scratch.

The phrase *on the other hand* functions as if it were one word, much like *however*. In fact, sequences like *upside down*, *by the way*, *the day before yesterday* have even been termed **polywords** (Lewis 1997). (From a pedagogical point of view, they might be treated similarly to long words which also demand precise memorization of several constituent parts.)

There is no established set or classification of multi-word items, and one random way of presenting them offered by Moon (1997: 44–47) includes compounds (*carpark*, *spin-dry*, *royal blue*), phrasal verbs (*write down*, *hang out*), idioms (*kick the bucket* – in a narrow sense, idioms have unitary meanings which cannot be predicted from individual word meanings), fixed phrases (*good morning*, *in fact*, similes and proverbs: *dry as a bone*, *it never rains but it pours*), and "prefabricated routines" also known as prefabs (*the thing/point is…*, *I'm a great believer in…*).

Lewis (1997: 9–11) observes that apart from fully fixed expressions (*by the way*; *not too bad, thanks*), people commonly use semi-fixed expressions – frames with slots that can be filled in a limited number of ways. These vary from short to long, and from almost fixed to very free. Examples include: minimal variation (*It's/That's not my fault.*), a simple slot (*Could you pass…, please?*), a sentence head that can be completed in many ways: *What was really interesting/surprising/annoying was…*). Furthermore, Lewis (1997: 30) distinguishes between strong collocations, which are linked so tightly that they behave like single words, and weak collocations (*nice day*, *good chance*), where each of two common words may co-occur with many other words.

There is no clear-cut distinction between what should and what should not be classified as a **multi-word item**, as opposed to other strings of words. Moon (1997: 44) presents three criteria characterizing multi-word items: institutionalization, fixedness, and non-compositionality. They describe the degree to which a given

word sequence (a) is conventionalized (institutionalized) and recurs in the language, (b) is a frozen sequence, and (c) has a holistic meaning which cannot be interpreted from individual words. However due to variability of forms of some multi-word items, the notion of fixedness need not always apply. This is illustrated with the examples collected from corpus analysis: *wash your dirty linen/laundry in public* (mainly British English), *air your dirty laundry/linen in public* (mainly American English), *do your dirty washing in public* (British English), *wash/air your dirty linen/laundry*, *wash/air your linen/laundry in public* (Moon 1997: 53).

2.5 The size of the lexicon

Interpreting and comparing estimates of lexicon size should be treated with caution for several reasons. First, we do not always know what is counted as a word. Secondly, there are major problems with assessing one's word knowledge. Thirdly, one word in the count need not be equal to another word in terms of its learning burden, understood as "the amount of effort required to learn it" (Nation 2001: 23).

When talking about a word, one might mean a type, a lexeme, or a word family. A lexeme contains a headword and possible inflected forms, while a word family consists of a headword, inflected forms and closely related derived forms. Changing a word count from word types to lexemes or from lexemes to word families can reduce the total number of items dramatically. As Table 1 shows, the entry for *number* includes an extra dozen derived forms, a considerable number of words to learn. The table also demonstrates that word families may exhibit great differences in size and in the concomitant learning burden.

Table 1 Word families for *number* and *drama* containing derived forms, with inflections omitted (adapted from *Longman Dictionary of Contemporary English* (2006) on CD-ROM)

noun	*adjective*	*verb*	*adverb*
□ **number**	□ innumerable	□ number	□ numerically
□ numeral	□ numerical	□ outnumber	
□ numeracy	□ numerous		
□ numerator	□ numerate ≠ innumerate		
□ innumeracy	□ numberless		
noun	*adjective*	*verb*	*adverb*
□ **drama**	□ dramatic	□ dramatize	□ dramatically
□ dramatist			

Results of word counts might depend, for example, on such decisions as whether or not words like *flat* and *flatten* or *police* and *policeman* belong to the same word families (Nation and Waring 1997: 8, 14). One thing that word family counts do not reflect is that – because learners keep extending their knowledge of affixes and learning more derived forms (members of the same word family) – during L2 acquisition word families tend to grow in volume, not just in number.

Various questions have been raised with reference to the size of lexicon: the language lexicon (e.g. How many words are there in the English language?), a native speaker's lexicon (e.g. How many words does a native speaker of English know?), an FL learner's lexicon (e.g. How many words does an FL learner need? How many words has one learned?), or a language dictionary (e.g. Which dictionary has more words?).

The most accessible way of estimating the number of all words in a given language is by consulting the most comprehensive dictionary, with the hope that it records all existing words of that language at a given point in time. The largest non-historical dictionary of English, *Webster's Third New International Dictionary,* is said to contain 114,000 word families, excluding proper names (Nation 2001: 6–9). Language is changing constantly as new words are created, some words become obsolete or take on new meanings, some word use is restricted to certain regional or social groups, and it would be unwise to expect any dictionary to contain complete word knowledge of a given language.

Nation (2001: 363) characterizes two basic methods of assessing people's vocabulary size: dictionary-based and corpus-based. (Corpora are large bodies of written texts or transcriptions of spoken texts which can be read and analysed by computers.) In the first, a dictionary presumably containing all the words known by the subjects is chosen, and a representative sample of words from it is selected and used to test L2 learners or native speakers. If 1% of the dictionary content was tested, the test result is multiplied by 100 to obtain a subject's vocabulary size. In the second method, a corpus of language (written, spoken, or both) is collected and words in it are counted. (This method has been used, for example, to estimate vocabulary used by Shakespeare or in US school texts.) The words are arranged as a frequency list and divided into groups – the thousand most frequent words, the second thousand most frequent words etc. Samples of these frequency groups may be used for testing language learners. Corpus-based frequency lists may be corpus-biased, e.g. some colloquial or taboo words may be infrequent in or missing from a corpus based only on written vocabulary from *TIME* magazine, as opposed to a corpus based on colloquial spoken language, which, in turn, may exhibit lack of some technical vocabulary.

As for the size of educated native speakers' lexicons, Aitchison (1987: 7) believes the number of words they know "is unlikely to be less than 50,000

and may be as high as 250,000." Nation (2001: 9) estimates that native speakers know about 20,000 word families, having added to their vocabulary an average of 1,000 word families a year in their early lives. 3,000–5,000 word families are necessary as a basis for comprehension (Nation and Waring 1997: 10).

Language learners' lexicon obviously grows throughout their L2 development. Reaching native speakers' level seems like a daunting task; however, it turns out that not all words are equally important. Knowing the most frequent words, we are able to understand large portions of texts. There have been different frequency lists compiled for English words, and among them Michael West's *A General Service List of English Words* from 1953 (based on written language) is a classic, still used. It consists of 2,000 word families, with a majority of content words (only 165 families are function words) (Nation 2001: 15). It is interesting that the first 1,000 most frequent words include as much as 84% of running words in conversation and 73% of academic text; however, the second 1,000 represent only 6% of conversation and 5% of academic text (Nation 2001: 17). Consequently, the 2,000 most frequent English words cover 90% of running words in conversation and 78% of academic texts. The coverage of academic texts will increase to 87%, if we extend the 2,000 word list by the *Academic Word List*, which comprises 570 word families not included in the most frequent 2,000 words, but appearing in a wide range of academic texts (science, commerce, law, and humanities) (Nation 2001: 17). To reach 95% coverage of academic texts (meaning that one in twenty words remains unknown), knowledge of 4,000 word families is necessary (Nation 2001: 147).

Referring to a study on reading comprehension of fiction text, Nation (2001: 147–150) suggests that 95% coverage is the absolute minimum, at which level most subjects did not gain adequate comprehension, however; while with 98% almost all learners achieved adequate comprehension. Therefore, coverage of less than 95% is recommended for intensive reading, 95–98% (one unknown word per 2–5 lines) for extensive reading for incidental vocabulary learning, and 99–100% coverage for intensive reading for fluency development and quick reading for pleasure.

A group of advanced students studying abroad in an L2 environment has been reported to exhibit an average vocabulary growth of as many as 2,500 words per year, similar to larger estimates of L1 growth in adolescence (Nation and Waring 1997: 8).

It is important to realize that growth of vocabulary knowledge can be greatly assisted by the use of an appropriate dictionary. Unfortunately, some learners get so used to their intermediate-level dictionaries that they find it virtually impossible to give them up for an advanced learner's dictionary once they have become more advanced. Part of the difficulty stems from the fact that some medium-sized monolingual dictionaries seem more appealing, having been

adapted to accommodate learners' L1 translations. *Oxford Wordpower* (2009) comprises 39% (72,000 items) of what *Oxford Advanced Learner's Dictionary* (2009, 7th edition) contains (183,500 words, phrases and meanings). The difference is even more visible in *Longman słownik współczesny* (2009), which contains 29% (66,500 items) of *Longman Dictionary of Contemporary English* (2009 edition – 230,000 words, phrases and meanings). The largest English-Polish dictionary, *Wielki słownik angielsko-polski PWN-Oxford* (2002), covers over 500,000 English *lexical items*. As rule of thumb, working with advanced texts requires at least advanced learner's dictionaries.

A major problem with word counts is that they tend to ignore multiple meanings of words. *LDCE* gives 9 meanings of a single noun entry for *bank*. Someone knowing just one meaning of *bank*, "a place where money is kept," is thought to know the word *bank*. Another person who knows all 9 meanings of the word bank is also described as knowing the word *bank*. Following *Wielki słownik angielsko-polski PWN-Oxford*, which has three entries for *bank*, we might tend to conclude that a person familiar with 3 meanings – "financial organization" (*bank1*), "land alongside a river or lake" (*bank2*), and "a row" (*bank3*) – in fact knows 3 words, all of which have the same spelling but completely different meanings. Looking at other different senses of the noun *bank1* (3 senses), *bank2* (8 senses), and *bank3* (2 senses) – 13 different senses altogether – one might wonder how many "*bank* words" there really are. Full knowledge of a polysemous word (one that has two or more different meanings) like *bank* requires much more than knowledge of a word with a single meaning, e.g. *giraffe*. The problem of word counts becomes even more complex when we note that the multiplicity of meanings concerns especially those words which are more frequent. *LDCE* lists 48 different senses of the verb *to go* (excluding phrasal verbs) and only 2 of the noun *ship* (both words belong to the first 1,000 most frequent words). How many senses of *go* do we need to know so that we can say we know the verb *go* – one? twelve? twenty-four? all forty-eight?

2.6 What does it mean to know a word?

Various studies have attempted to address the problem of what it means to know a word, with different degrees of overlap (e.g. Aitchison 1987, Nation 2001, Richards 1976, Singleton 1999). Rather than compare them, we are going to present one comprehensive approach, formulated with direct pedagogical implications in mind. Nation (2001: 26) considers that "[a]t the most general level, knowing a word involves form, meaning and use." He discusses each of these three components in more detail with regard to aspects presented in

Table 2. He also stresses that each of those aspects has two sides, relating to *receptive* and *productive* knowledge.

Table 2 Aspects of word knowledge (adapted from Nation 2001: 27)

Word knowledge		
form	meaning	use
• spoken form	• connecting form and meaning	• grammatical functions
• written form	• concepts and referents	• collocations
• word parts	• associations	• constraints on use

2.6.1 Spoken form

We know the spoken form of a word receptively if we can recognize the word when we hear it, and we know it productively if we can pronounce it to express its meaning. Pronounceability of a word is largely determined by the degree of similarity between L1 and L2 sound systems and the ways sounds in both languages combine. Because of problems with the English sound system, some learners might say that they *fink* ("think") or like to *sink* (or *sin*) ("sing") in the shower. A recurring mistake made by some students planning on writing their M.A. theses involves mispronouncing /ˈθiːsɪs/ in such a way that what they are actually saying is, for example, *I want to devote my faeces to modern American poets*. Some pronunciation problems are related to orthography; for example, in Polish each vowel letter has a corresponding vowel sound, while in English there are various ways of pronouncing the same vowel letter (cf. *got*, *done*, *so*, *frown*, *bought*, *anorak*, *to*).

2.6.2 Written form

The irregularity and unpredictability of the English spelling system account for a large number of problems with spelling words. One might know the pronunciation of a word but not be able to write it correctly and vice versa. Interestingly, the development of electronic communication seems to have yielded more tolerance for messages that are incorrect linguistically but efficient from a communicative and pragmatic perspective, but at the same time the importance of exact spelling has come to the fore – accessing a desired address or using a password demands that we use every single digit with utmost precision. Also, in business, the change of one letter (cf. *Panasonic/Panasonix*) might have fundamental consequences worth millions of dollars.

2.6.3 Word parts

Knowledge of word parts includes, for example, perceiving differences between regularities underlying the meaning of *-ship* in words like *friendship* and *battleship*. Understanding the meaning of some roots or prefixes (e.g. *hypo-* and *hyper-*) enables us to deduce the meaning of words we have not seen before (e.g. *hyposensitive* and *hypersensitive*). However, deceptively transparent words (Laufer 1997: 146) often cause problems, because we do not realize that they are unfamiliar to us. Sometimes learners assume incorrectly that the nouns *undercoat* and *understudy* refer to a garment and a student, rather than a layer of paint and an actor.

2.6.4 Connecting form and meaning

One may be familiar with the form of a word – written or spoken (e.g. *ignominiously* or /ʃəˈnænɪgənz) – but not know the meaning it corresponds to. One may also know the word form, e.g. *thyroid*, and know the concept behind it, i.e. be roughly familiar with human anatomy, but still not be able to connect the form *thyroid* with its proper meaning. Connecting form and the proper meaning is an essential part of knowing a word. The stronger the link between form and meaning, the easier it is to understand the meaning of a word and retrieve the correct form for the meaning we want to express, and every time we retrieve from memory the form or the meaning of a word, the connection becomes stronger.

2.6.5 Concepts and referents

As already mentioned, many words, especially high-frequency words, have multiple meanings, which is not reflected in word counts but definitely increases the learning burden of a word, much like homonymy does. **Homonyms** are words which have the same form but differ in meaning, and they ought to be counted as separate words. Identity of form may refer to both spelling and pronunciation, as in *ear* (a body part or part of a plant such as corn), or just one medium – then we talk about two types of homonyms: **homographs**, which have the same spelling but different meaning (e.g. the noun *tear*/the verb *tear*) and **homophones**, which are pronounced in the same way but differ in meaning (e.g. *nose/knows*).

There are two general approaches describing how our mind processes words with multiple meanings. According to one, each sense of a word is stored separately and every time we process a word, we choose the proper meaning in **sense selection**. The other stipulates that a word has one core meaning,

underlying all specific meanings, and we search for the correct meaning through **reference specification** (Nation 2001: 50).

2.6.6 Associations

An essential part of lexical development involves building relationships between words such as, for example, synonymy, antonymy, hyponymy, entailment or troponymy. **Synonyms** are words with the same meaning; however, total synonymy, where words are used correctly with identical meanings in all contexts and with the same connotations, is rare. Words are more commonly synonymous in some but not all contexts, or have different connotations. For example, *range/selection/choice* are synonyms in the phrase *a nice range/selection/choice of flowers*, but not in the phrase *his range/selection/choice of knowledge*, where one of the possible synonyms of *range* might be *breadth* (Crystal 2008: 470). **Antonyms** have opposite meanings, e.g. *big/small*. Antonymy can be meaning-based (*good/bad*, *interesting/boring*) or form-based (*interesting/uninteresting*).

Nouns can be arranged in hierarchies consisting of several levels, e.g. *animal – mammal – dog – puppy*, and the relationship between members of such a hierarchy is called **hyponymy** (the prefix *hypo-* means "under" – *hypothermia* is a medical condition where body temperature is much below normal). In the case of a pair like *mammal/dog*, *dog* is a hyponym of *mammal* and *mammal* is the hypernym.

Still another type of relationship is called **troponymy**. It concerns verbs and is related to the way in which an activity is performed; for example, *march* is a troponym of *walk*. A relationship more general than troponomy is that of entailment. *March* entails *walk* and *snore* entails *sleep* (but it is not a troponym of *sleep*).

2.6.7 Grammatical functions

Sometimes learners might be familiar with the semantic content of words, but use the wrong part of speech; for example, Polish students often use *relax* as a noun and have problems with an adjective and adverb *live* (*a live rhino*; *they played live*). The choice of words, especially verbs, has an effect on the structure of a sentence. Differences in grammatical patterns in L1 and L2 account for the following mistakes: **He suggested me this* (the English counterpart of *proponować* + *komuś* + *coś* should be *suggest* + something + *to* someone) or **Jadłem dużo owoc* (the adverb *dużo* requires the genitive case and *owoc* needs to be used in the plural (*dużo owoców*)). The choice of the the verb *influence* determines that we say *He influenced your decision* (without a preposition), but if we choose the noun *influence*, it has to be followed by a preposition:

He had influence over your decision. Choosing an adjective that can be used only predicatively, like *afraid*, means it cannot be used before a noun (*I saw people who were afraid. *I saw afraid people.*)

2.6.8 Collocations

The ways in which words co-occur and form collocations are often language-specific and cause considerable learning difficulty. Polish learners who know that the English equivalent of *tarty ser* is *grated cheese* have no way of predicting that the counterpart of *bułka tarta* ("grated roll") is *breadcrumbs*. Similarly, knowing that *blustery* means "windy" – as in *blustery day* ("wietrzny dzień") – is not enough to predict the collocation *blustery wind* ("porywisty wiatr"). There is a certain degree of expectancy between elements of a collocation which is recognized by native speakers, but because there is an indefinite number such word combinations, getting close to a native speaker's intuition in respect of collocation knowledge is one of the most difficult goals to achieve in L2 acquisition.

Learners are not always aware that a given word might change its meaning, depending on the context in which it appears and the word(s) it collocates with. Compare the meaning of the word *odd* in the following: *the odd thing is* ("the strange thing is"), *odd socks* (socks that do not match), *the odd drink* (an occasional drink). An amusing failure to distinguish context-sensitive meaning is illustrated by an English-Polish online translation of an article found on the Internet, where *canine extraction* ("canine tooth extraction") was translated as "pulling out a dog."

2.6.9 Constraints on use

English speakers have to discover how to use "you" when learning Polish, where the pronoun *ty/wy* (with the concomitant verb forms) indicates a degree of speaker-listener familiarity and *Pan*/*Pani* a degree of social or psychological distance, and they have to find how this relationship is defined in different social groups, between colleagues of different ranks, between people of the same age, when talking to much younger people, etc. Polish speakers need to be aware, for example, that the word *pussy* might sound neutral in British English when talking about a cat to a child, but otherwise be offensive when talking to a woman.

2.6.10 Productive and receptive word knowledge

The more words learners know, the greater the difference between active and passive vocabulary sizes. The difference is described as more visible with

low-frequency words, high-frequency vocabulary tending to be known both receptively and productively (Nation 2001: 370–371). Laufer's (1998) study on vocabulary growth shows that the difference between active and passive vocabulary size is bigger for ESL learners than for EFL learners, which seems due to input differences.

2.6.11 Testing word knowledge

Until recently, vocabulary testing has concentrated on assessing the number of words learners know. In its crudest form, a vocabulary size test is a checklist (Meara 1996) – subjects are given a list of words and they need to check those they "know". However, to provide a more adequate account of one's vocabulary, we should include (a) vocabulary size or breadth, i.e. the number of words known ("how many"), (b) depth of knowledge, i.e. the amount of knowledge of each word ("how well"), (c) automaticity, i.e. how quickly this knowledge can be used, and (d) organization of the lexicon (Schmitt 1997: 104, 326).

An attempt to measure vocabulary depth and to give credit for learners' **partial knowledge** of words might be illustrated by Wesche and Paribakht's (1996) Vocabulary Knowledge Scale, in which learners respond to the following statements during an interview:

1. I haven't seen this word before.
2. I have seen this word before, but I don't know what it means.
3. I have seen this word before and I think it means…
4. I know this word. It means…
5. I can use this word in a sentence.

No vocabulary depth test measuring all aspects of word knowledge discussed above has been developed to date, and attempts to include a large variety of aspects have resulted in the tests becoming so time-consuming as to become impractical.

A technique to assess the organization of the mental lexicon is an **association chains** task, where the subject is given a stimulus word and a target word and is asked to complete the chain of associations. For example, *sea – – – butterfly* has been completed in many different ways, including *sea – horse – horsefly – butterfly* and *sea – fish – fly – butterfly* (Meara 1996).

2.7 Mental lexicon

A native speaker needs about 200 ms (1/5 of a second) to recognize a word in his or her language, and about half a second to reject a non-word (i.e. to state that a given sequence of sounds or letters does not constitute a word in

a given language) (Aitchison 1987: 7). This speed is astounding. If one tried to run through a lexicon containing 50,000 running words at the rate of 50 words per second, that would take more than 16 minutes. The actual speed of lexical decisions, as well as the fact that words are recognized even before the whole word is heard, indicate that our mental lexicon has some kind of organization system. Our **mental lexicon** is the mental representation of our knowledge of vocabulary, stored in our long-term memory. We use this lexicon in comprehension by looking up words in our memory and in production by retrieving the words we need. Two important questions arise here: How is the mental lexicon organized? (How do we store words in it?) and how do we retrieve words from it? Suffice it to say at this point that there is evidence indicating that words are stored and retrieved on the basis of their phonological and semantic properties.

The mental lexicon is not just the sum of its parts, but also connections between them. Meara (1996) emphasizes the need to analyse the lexicon in terms of its size and its organization/structure, understood as the degree of connectivity between lexical items.

Analyses of the mental lexicon often use terms such as **processing**, **storage/retention**, and **access/retrieval**. Processing refers to "cognitive operations which take place during language production and comprehension. Any level of language can be considered in processing terms ('lexical processing', 'phonological processing', etc.)" (Crystal 2008: 388). Lexical storage refers to "the way words are retained and made available for use by the brain during language production and comprehension. It forms part of a theory of language processing" (Crystal 2008: 278). Lexical access is "the sequence of processing stages which have to be postulated to explain how speakers retrieve words from their mental lexicon" (Crystal 2008: 276).

We gain insight into the functioning of the mental lexicon by using different kinds of evidence, e.g. slips of the tongue, tip-of-the-tongue (TOT) phenomena, and aphasia studies (involving people with brain damage), as well as word-association, reaction time, and priming experiments. Neurolinguistic methods include measuring electric activity of brain cells by means of EEG (electroencephalography) and detecting areas with increased blood flow (supposed to indicate increased information processing) by means of fMRI (functional magnetic resonance imaging) or PET (positron emission tomography). Unfortunately, "literature on bilingual neuroimaging is fraught with inconsistent results and conflicting interpretations" (Paradis (2004: 184).

We know from the tip-of-the-tongue phenomenon, where we are almost but not quite able to recall the word we are searching for, that the words that come to our mind are likely to have the same stress pattern and the same initial and final sound or letter as the target word.

Reaction time experiments measure the amount of time it takes to complete a task, for example, to state whether a word that a subject sees or hears is a word or a non-word. This kind of experiment often includes **priming**, which involves the influence of a previous stimulus, called a **prime**, on another stimulus, called a **target**. Let us assume that a subject sees the word *driver*, states that it is a word, then sees *nurse*, and again recognizes it as a word. It turns out that if the first word (the prime) is not *driver* but *doctor*, it takes less time to recognize *nurse* (the target) as a word than it does when *nurse* is preceded by *driver*. In this semantic priming experiment, the prime facilitates the processing of the target if both the prime and the target are semantically related.

2.8 Models of lexical processing

A model of the mental lexicon is not a replica of it, but a theoretical simplified construct involving a large amount of guesswork. Such a model aims to reflect the principles and workings of a mental lexicon, but even if the model accounts for observable outcomes, it does not necessarily mean that it adequately describes the underlying processes (Aitchison 1987: 28–29).

Singleton (1999: 84) makes a distinction between direct (one-stage) and indirect (two-stage) models of the mental lexicon. In indirect models we access a word in two stages, involving a search procedure and a retrieval procedure (similar to the way we obtain a book from a library). Direct models have a one-stage access (similar to word processing software, where every item stored has a name and in order to access the one we need, we only need to type in as many letters as are necessary to distinguish this item from all other items). Direct models include the logogen model and the cohort model, while examples of indirect models include the lexical search model and the connectionism model.

2.8.1 The logogen model

An early version of the logogen model (Morton 1970, cited in Singleton 1999: 86) contains three components: the logogen system (a logogen is a neural unit, part of the nervous system; there is one logogen for each lexical item in one's mental lexicon), the cognitive system (containing semantic information), and the response buffer (responsible for word production, both spoken and written). The logogen system collects acoustic and visual evidence from language input, as well as semantic evidence from the cognitive system. All this information excites many logogens, but the model stipulates that only one logogen "fires" when activation reaches a critical threshold, and the proper word is selected.

(In order for this to happen, there are actually two thresholds – one controls access to the cognitive system and one controls access to the response buffer.) Each logogen has a particular activation level, which depends on how often it has been activated. One of the problems with this model involved specifying activation levels for individual logogens and identifying critical thresholds. The solution to this problem is offered by the cohort model, presented below.

2.8.2 The cohort model

The cohort model (Marslen-Wilson and Tyler 1980) assumes that when we hear a word, the acoustic information activates word detectors. They are activated the moment we perceive the beginning of a word. For example, when someone starts saying the word *parliament* and we hear /pɑ:-/, all word detectors for words beginning with /pɑ:-/ are activated. This whole set of words which we access (activate) in our mental lexicon before the single target word is selected is called the **word-initial cohort**, and in our example, it would include, for instance, *par*k, *par*ka, *par*lance, *par*ley, *par*liament, *par*lor, *par*lous, *par*sley, *par*tial etc. When the next sound is uttered and we hear /pɑ:l-/, those words which do not match the acoustic input remove themselves from the the word-initial cohort, which is then narrowed down to: *parl*ance, *parl*ey, *parl*iament, *parl*or, *parl*ous, etc. Next we hear /ˈpɑ:lə-/, which limits the range of words to *parla*nce, *parlia*ment, *parlo*r, and *parlou*s. As soon as /m/ is heard, the word *parliament* is selected, because there is no other English word that begins with /ˈpɑ:ləm-/ (the word *parliamentary* is not likely due to a different stress pattern). /ˈpɑ:ləm-/ is the **uniqueness point** for parliament. Each word has its uniqueness or recognition point, the exact point when the word is recognized.

The model recognizes the role of context in lexical processing and stipulates that after a word-initial cohort has been selected, contextual factors come into play. In a word-monitoring study, subjects presented with auditory material were supposed to press a button as soon as they heard a word they had been given before. The listening material included (1) normal, coherent sentences, (2) sentences correct syntactically, but anomalous semantically, and (3) random strings of words. As the support provided by context decreases, the subjects' mean reaction times increase (273 ms, 331 ms and 358 ms, respectively).

The cohort model allows one to specify points of recognition of non-words (a **non-word** is a sequence of phonemes or letters which is not an existing word in a given language, e.g. *pychology*, *trrukosf* or *ptak* are non-words in English). For English, the initial sound sequence /pt/ will mark the uniqueness point for non-words starting with /p/, because no English word begins with the sound sequence /pt/. Experimental evidence supporting the theory underlying the model shows "that the time taken to recognize non-words will be shorter where

recognition points come early in words and longer where recognition points come late" (Singleton 1999: 93).

Although the cohort model deals with speech recognition and analyses input on the phonological level, it has important implications for the retrieval of lexical items with regard to orthography, outside the realm of the mental lexicon. A number of electronic dictionaries specify the beginning of word-initial cohorts (e.g. CD dictionaries published by Longman, CUP, OUP, or Collins). The function of identifying a uniqueness point and automatically providing the rest of a word which is being selected has been adopted by Microsoft Word (*autocomplete* option), Google, and some cell phone text-messaging programs.

2.8.3 The Lexical Search Model

The Lexical Search Model (Forster 1976) is an indirect model. Singleton (1999: 99ff) discusses the model beginning with an analogy to the two stages we go through when finding a book in a library. First, depending on the kind of information we have about the book we want (author, title, or just subject), we select an appropriate book catalogue and search it to find the appropriate shelf mark. Then, knowing the exact location of the book, we go to a specific shelf and get the volume. The two-stage processing in the lexical search model involves peripheral **access files**, which are connected by means of pointers to the unitary master file. In receptive language processing, we are thought to go to a **phonological access file** (for the words we hear) or an **orthographic access file** (for the words we read), while if we produce language on the basis of meaning intentions, we go to a **syntactic/semantic access file**. Each access file lists words according to phonological, orthographic, and syntactico-semantic criteria, respectively, together with pointers, which (just like shelf marks) point to the exact location of a complete entry in the master file. The model does not assume any connectivity between different access files, but it allows for some cross-referencing in the master file between closely related words. Experimental results, however, do not seem fully to support Forster's theory.

2.8.4 Levelt's blueprint for the speaker

A well-known language processing model is Levelt's Blueprint for the Speaker (1989). Although it is not exclusively a model of lexical processing, the lexicon plays a central part in it. Levelt distinguishes between **declarative knowledge** ("knowing that"), which can be examined reflectively, through introspection, and **procedural knowledge** ("knowing how"), which cannot be accessed through conscious thought. The lexicon component in Levelt's model contains declarative knowledge and consists of lemmas (semantico-grammatical

knowledge) and forms/lexemes (morphophonological knowledge). Each lemma contains the word's meaning, conceptual argument structure, syntactic category (part of speech), grammatical profile, and possibly information about constraints such as style and appropriatness, which match particular contexts. The lemma is linked by means of a pointer with the morphophonological form of the word.

The procedural components of Levelt's model work as follows. After the Conceptualizer has generated a message, the Formulator gives the pre-verbal message a syntactic and phonological shape, yielding internal speech, which the Articulator transforms into overt speech. The remaining components of the model are responsible for self-monitoring (analysing produced speech and adjusting further production).

Levelt (1989: 181) considers speech production procedures lexically-driven. The pre-verbal message triggers lexical items into action, and these, in turn, trigger grammatical, morphological, and phonological encoding in the Formulator. In other words, the grammar and phonology of generated utterances are determined by characteristics of lexical items.

In his analysis of the lexicon, Levelt (1989: 183) distinguishes between relations "within entries" and "between entries". He treats inflected forms as part of the same entry, and derived forms as separate entries. Furthermore, he divides relationships between entries into intrinsic relations, based on meaning, grammar, morphology, and phonology, and associative relations, based on frequent collocations (e.g. *green* and *grass* or *thunder* and *lightning*). As for intrinsic relations, words are related semantically when they belong to the same lexical set (parts of a car or names of the months) or constitute synonyms or antonyms. Grammatically related words are the same part of speech of perform the same grammatical function. Members of the same word family are related morphologically, through derivation. Phonological relations might involve words beginning with the same sound.

2.8.5 Connectionism

When brain cells are active, they send signals to other neurons. Excitatory signals cause arousal, while inhibitory signals cause suppression. Thus, in a network of interconnected units, some units are reinforced as a result of arousal, and the connections with others are lost as a result of inhibition (Aitchison 1992: 31). The approach which explains mental processing by analogy with brain activity and neural networks is called **connectionism** (or parallel distributed processing). Its most striking feature is that it represents knowledge not in terms of entities which are stored in our mind, retrieved, and combined according to patterns or rules (which are also stored), but in terms of connection strength (Singleton

2000: 179). Connectionism uses the concept of spreading activation and is related to two other types of models.

2.8.6 Spreading/interactive activation models

In **spreading activation** models, the mental lexicon consists of interconnected nodes, where the arousal of some nodes spreads and excites more and more other nodes to which they are connected. **Interactive activation** models posit that the activation flows not just in one direction between activated nodes, but forwards and backwards. Connectionism, unlike the other two models, postulates inhibitory signals, not just excitatory ones.

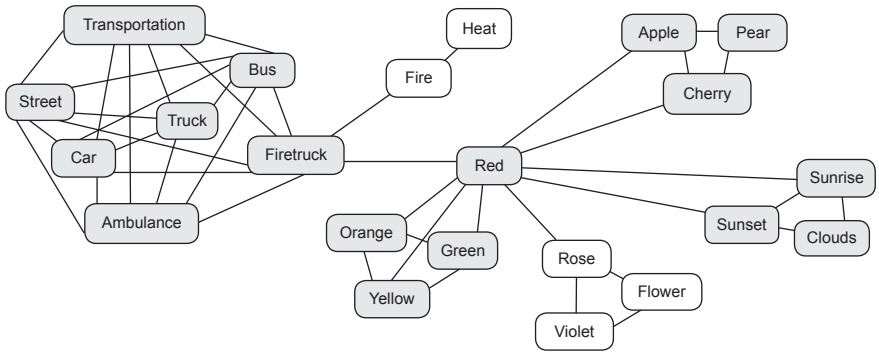

Figure 1 A Spreading Activation Model (Randall 2007: 115, cited in Kersten 2010: 19)

2.8 Bilingual lexicon – integration or separation?

Unfortunately, we do not know the exact structure of the L1 mental lexicon, and the situation becomes even more complicated if we include the lexicon of L2. The basic question asked here is whether the bilingual lexicon is a unified whole, or whether it contains separate language lexicons. There are studies supporting both **integration** and **separation**. For example, interference errors are incorrect forms originating in L1 (*Może być!* → **Can be!*). On the other hand, bilingual studies on aphasia and language loss show that language disorders and language recovery may affect one language, but not the other (Singleton 2003). It might be assumed that different types and degrees of **interconnectivity** between the L1 and L2 lexicons "will make some parts of our mental lexicon more integrated and others separate" (Gabryś-Barker 2005: 50).

Kroll (1993, cited in Kersten 2010: 28) suggests a model stipulating that bilinguals in both L1 and L2 rely on the same conceptual store, and that

the strength of lexical and conceptual links changes with proficiency level and age of acquisition (Figure 2). Initially, the concept behind a new L2 word is arrived at via a corresponding L1 word. With time, direct links between L2 words and their underlying concepts are established.

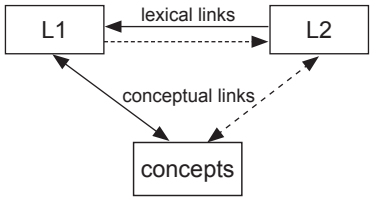

Figure 2 Concepts and L2 acquisition (Kroll 1993: 54, cited in Kersten 2010: 28)

Some of the drawbacks of studies on word knowledge and lexical storage and access include using single words in isolation, in situations different from authentic communication, and disregarding the context of language use, which can strongly influence the meaning of a word, its connotations, appropriatness of use, or interpretability (Read 1997: 319).

2.10 Lexical development in L1 acquisition

Children acquiring their first language face a formidable task: they have to isolate word units from the continuous speech stream they hear, they have to create potential meanings, and they need to map the meanings onto the forms (Clark 1993: 43). We communicate meanings via concepts which are mental representations of referents (objects, persons, places, attributes, actions etc. in the "real world") (Singleton 2000: 65).

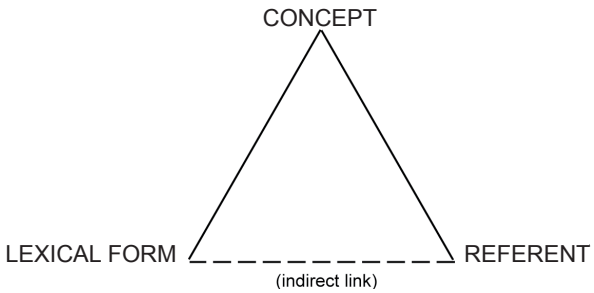

Figure 3 Associations between the concept, the lexical form, and the referent (Singleton 1999: 30)

Early child language development includes four important stages, summarized in Table 3 (based on Singleton 1999: 61). Cooing and babbling are pre-speech stages, followed by so-called true speech – production of one-word meaningful utterances.

Table 3 Milestones of early child-language development (based on Singleton 1999: 61)

Stage	Child's age when this stage begins	Child's output
Cooing	1–4 months	vowel-quality vocalizations
Babbling	4–8 months	combinations of both vowel-like and consonant-like sounds
One-word utterance stage	end of the first year	meaningful single-word utterances
Two-word utterance stage	18–24 months	producing words within the same tone group

After a child has produced its first words, lexical development takes place in three phases. In Phase 1, lexical development is very slow, and it might even take half a year for a child to produce the first dozen words. During Phase 2, a **vocabulary explosion** – or "vocabulary spurt" (Clark 1993) – phase, the child acquires a large number of words very quickly. This usually happens after a child has already acquired about 30 or more words. Phase 3 is characterized by consolidation, revision, and reorganization of lexical knowledge. It starts in pre-school years and continues into adulthood (Singleton 2000: 167).

Research on concept formation proves the existence of concepts in a child's mind at the pre-speech stage. An experiment was carried out based on a finding that infants exposed to two stimuli – one they have already seen and one they have not – will look longer at the new stimulus. Infants aged 3–4 months were familiarized with distorted versions of a simple geometrical shape (square, triangle, or diamond). Afterwards, they were shown undistorted shapes – a version of the shape already seen and a new one. The infants looked longer at the novel shapes, which indicates that they have created **prototypical concepts** of the shapes correctly, even though they had been exposed to distorted versions of those shapes (Singleton 1999: 55).

We know that concepts exist in a child's mind, but we do not know where they come from. Chomskyans and **nativists** maintain that they are innate and that a child's lexical, or linguistic for that matter, development is possible due to certain innate properties and special mechanisms enabling any human being to acquire a language. Some researchers do not entirely support this view and point to other, external facilitating factors. For example, adults tune a child's

input (they modify the language directed to children) by making their utterances shorter, speaking more slowly, and repeating key words, which helps a child to extract words from an adult's speech. Child-directed speech thus modified is referred to as **motherese** or **caregiver's speech**. An inherent feature of motherese is ostensive definition: saying a word and simultaneously pointing to an object being named. Doing this, an adult isolates a word for the child and facilitates mapping its meaning onto the form.

Early word meanings among children aged one or two are characterized by under-extension and over-extension. For example, a child might under-extend the word *dog*, using it for most prototypical dogs, excluding a Pekingese or a Chihuahua (Clark 1993: 34), or over-extend *dog* to goats, sheep, calves or cows.

Lexical expansion during the vocabulary spurt (Phase 2) indicates a child's discovery that everything around has a name (naming insight) and vocabulary growth is aided by fast mapping – great ease in acquiring new words even after very little exposure. Clark (1993: 28, 31) observes that even though children might know not more than 50 or 100 words, they use words from different domains or semantic fields, such as people, food, animals, vehicles, body parts, or clothes. Initially, however, each domain might be represented by just one or two words, e.g. *mama*, *milk*, *dog*, *car*, *nose*, *shoe*, respectively. As the children's vocabulary expands, so does the size of each domain; thus, *duck*, *mouse*, and *cat* might be added to the animal domain, followed by *horse*, *fish*, *goose*, etc. Furthermore, new domains are created.

In Phase 3, important reorganization of lexical items takes place. Language users approach word relationships in a more general and abstract way and develop hierarchical levels in word meanings. Thus, a child who has learned that *tulips* and *roses* are *flowers* and *oaks* and *elms* are *trees*, will cluster both semantic domains into one, using a superordinate term, *plants*. Similarly, after *baby*, *daddy*, etc. have been classified as *people*, and *dog*, *cat*, *tiger* as *animals*, a still more general superordinate term might be used further up the hierarchy: *plants*, *people*, and *animals* are members of a broader domain which might be labeled *things that are alive* (Singleton 2000). In this way, extralinguistic reality starts to be classified through superordinatness and hyponymy.

Another important aspect of the growth of mental lexicon is the **syntagmatic-paradigmatic shift**, taking place between the age of five and ten. Free association tests show that younger children respond with syntagmatic associations (where a response is a different part of speech than a stimulus, e.g. *red – apples*, *table – eat*), which are largely replaced by paradigmatic associations (where a stimulus and a response are the same part of speech, e.g. *red – green*, *table – chair*), predominating in older children and adults.

2.11 Some aspects of L2 lexical acquisition

Unless both L1 and L2 are acquired simultaneously (for example, when each parent speaks a different language with a child), L2 acquisition usually takes place when we are at a different level in terms of cognitive, linguistic, or psychosocial development, and we have already developed a phonological and a conceptual system with semantic hierarchies, possibly with a certain degree of literacy. We do not go through cooing and babbling stages in L2, but we have our L1 lexical knowledge as a reference point or backdrop, whether we want this or or not.

Isolation of a spoken word form might be problematic even for advanced Polish learners of English, partly due to different vowel and word stress systems and vowel reduction in English. Some learners report that they find it easier to establish and remember the phonological form of a new word if they see its written form first. Consequently, there are times when retrieval of the phonological form is not primary but follows retrieval of the written form of a word. Word form recognition is important inasmuch as learners might confuse meanings of different words that are similar in form or take an unknown word for a known one.

Phonological short-term memory (measured by the length of a sequence of unknown words or non-words that a subject can repeat in the correct order) is a reliable predictor of one's ability to acquire vocabulary later on, in both L1 and L2 (Ellis 1997, Singleton 1999).

Arabski (1996b: 78) notes a shift in emphasis in acquisition studies of L2 word phonological structure, from comparisons of L1 and L2 forms to studying retention of L2 items.

A major factor distinguishing naturalistic L2 acquisition from L2 learning in formal settings is the amount and kind of input, and for many learners this might be limited to classroom activities. Acquisition of lexis is enhanced by encountering new words in natural situations and is assumed to involve guessing word meaning from context. Extensive reading is strongly encouraged to ensure incidental vocabulary acquisition. At the early stages, however, we face what is called the **beginner's paradox** (Kersten 2010: 75) – learners are simply unable to learn from context, because they do not know enough of the words that constitute the context. Research has shown that in order to learn efficiently from context, using unsimplified texts, a vocabulary of about 3,000 words is required, and that number secures coverage of at least 95 per cent of a text (Nation and Waring 1997: 11). Nation and Waring (1997: 11) recommend that learners be taught the 3,000 most frequent words as soon as possible (even in a non-contextual manner, using word-cards), and then be given help in developing strategies to understand and acquire low-frequency words. An important way

of acquiring large numbers of new words at that later stage is through extensive listening and reading.

It is important for teachers to be aware of both the amount and kind of input their learners receive, as well to recognize their areas of difficulty. Schmitt (1977: 231) discusses a study in which four groups of teachers of English were asked to predict lexical difficulty of words. On the whole, non-native speakers of English (teachers whose L1 was the same as their students') did better than English native-speakers. Furthermore, students' knowledge of the target words was best predicted by non-native, inexperienced teachers, who found it easier to see the text through the learners' eyes.

> The English-speaking trainee teachers demonstrated an astonishing lack of awareness of word difficulty and, ironically, the experienced group of native speakers with postgraduate qualifications argued eloquently from the literature in support of their inaccurate predictions. (Schmitt 1997: 232)

Motherese, facilitating L1 acquisition, has its counterparts in both naturalistic L2 acquisition and more formal settings. **Foreigner talk** is the kind of language native speakers use while talking to foreigners, employing shorter sentences, speaking more slowly, and limiting their vocabulary. Similar features are demonstrated in **teacher talk**, where words have more general meanings and are less idiomatic.

Nation (2001: 63–65) describes three processes that lead to a word being remembered: noticing, retrieval, and creative (generative) use. Important aspects of noticing, or giving attention to a word, are interest, motivation and decontextualization, which here means regarding a word as part of a language system rather than as part of a message. Vocabulary is learned better if it involves negotiation of meaning (learners might be involved in negotiation themselves or they might also watch others negotiate). Retrieval does not occur when learners are presented with the form and meaning at the same time, but when only form is presented and they need to retrieve its meaning while listening or reading (receptive retrieval), or when they speak or write and want to communicate a certain meaning and need to retrieve its spoken or written form (productive retrieval). Every time a word is retrieved, the link between its form and meaning is strengthened, which facilitates future retrieval, provided the next retrieval takes place soon enough so that the word has not faded from one's memory. In generative processing, learners – while listening or reading – come across a word used in a different way than previously (receptive generation), or they use a word in a new way in a new context (productive generation).

Are morphological relationships reflected in the mental lexicon? Studies show that the speed of recognition of words is affected by inflectional and

derivational relationships, which suggests that "inflected and derived form are stored under the same entry or are linked to each other in the mental lexicon" (Nation 2001: 269).

Failure to connect word form and meaning properly might result, for example, in **underdifferentiation** errors, when two different L2 words have the same translation equivalent in L1 (e.g. *borrow* and *lend* have one translation equivalent in Polish – *pożyczyć*). Accounting for lexical development in L2, Arabski (1979: 34) uses the notion of the **primary counterpart**, defined as the L2 "equivalent which in the process of foreign language learning is acquired to render the common meaning of a given L1 lexical item." Primacy is ascribed by statistics. The Polish verb *mieć* is often translated into English as *have*, and as a result learners tend to acquire *have* as the primary counterpart of *mieć*, which results in mistakes such as **I have 19 years*. With time learners find that *mieć* has another English counterpart, *be*, and start producing correct forms, like *I am 19*. With some words the list of their counterparts may be rather long, which makes learning a given word more complex and time consuming.

False friends, words in two languages which are similar in form but which express different meanings, are often sources of errors (cf. Rusiecki 2002). By way of illustration, a student fervently described to an examination board a research project he was planning to carry out, based on *his own autopsy* (Polish *autopsja* means "personal observation or experience").

While early stages of L2 acquisition are marked by substantial reliance on L1 and interference errors, in later stages learners have found that similarity between L1 and L2 may be deceptive and as a result unknown L2 multi-word items (e.g. idioms or phrasal verbs) which are seen as too similar to L1 forms are treated with suspicion and might be avoided (Hulstijn and Marchena 1989, Kellermann 1977, Kellermann 1978) or even rejected as incorrect despite being perfectly acceptable L2 forms, only to be accepted again at a still higher proficiency level (Kellerman 1985).

Verspoor and Lowie (2003) found that the acquisition of polysemous words was greatly facilitated if learners were provided with a core sense underlying different meanings of a polysemous word. Those who were given the core meaning were better not only at guessing an abstract sense of a polysemous word, but also at remembering the word – their long-term retention scores were significantly higher.

Defining a core meaning successfully might not be always possible, and due to the deceptive nature of polysemous words, learners are often unaware that they do not know the meaning of a word being used. For example, someone who comes to an airport and is asked to check their baggage, and is then asked again why they have not done so yet, may be utterly confused, convinced that *checking*

baggage means "examining" it, and unaware that under the circumstances it means "leaving" it at a designated place.

Word associations studies indicate that early learners of L2 tend to produce syntagmatic and clang associations (Meara 1978, 1982) (the latter concerns words related in phonological terms only), while more proficient groups are apt to use paradigmatic associations (Söderman 1993), which is reminiscent of the syntagmatic-paradigmatic shift in L1 acquisition that has been described above. Word familiarity is a significant factor in word association, as Söderman's study suggests that less frequent words were more likely to elicit syntagmatic reactions.

Among grammatical categories, nouns are said to be the easiest to learn; they are followed by verbs and adjectives, adverbs being the most difficult. The effect of part of speech is supposed to be greater in lower levels of proficiency. There is some disagreement as to whether concrete nouns are always easier than abstract ones (Laufer 1997: 148–149).

Lexical density as defined by Ure (1971) reflects the proportion of content to function words. It is measured by dividing the number of content word tokens by the number of all running words and multiplying by 100%. Written texts tend to have a higher lexical density than spoken ones. In his study of written compositions of L2 learners, Arabski (1979; 1996a: 141) found that lexical density increased with proficiency levels from 47% (beginners) to 69% (intermediate) to 77% (advanced). (Some researchers (e.g. Crystal 2008: 276) define lexical density as synonymous with type/token ratio, arrived at by dividing the number of word types (different words) by the number of word tokens (all running words) in a text and multiplying the result by 100. The higher the ratio, the richer the vocabulary, and the more difficult the text is likely to be.)

Reaction time studies show that some frequent collocations are treated as lexical units and processed faster than if they were recreated from individual words (Nation 2001: 335). This indicates that words are stored individually and also as parts of larger chunks retrieved without being analysed into individual units. A large proportion of utterances consists of familiar word sequences, which are not entirely new creations. Native-like fluency incorporates collocational knowledge. Staying in a target-language-speaking country increases a learner's fluency due to a greater number of stored memorized sequences, as revealed by measuring the length of syllable sequences uttered without a pause (Nation 2001: 323).

2.12 Final remarks

Despite the large amount of research that has been carried out, acquisition of vocabulary is far from being fully understood and we are lacking an overall

theory of lexical acquisition (Meara 1997). Due to the lack of an exact definition of a lexical unit and lack of satisfactory instruments, we cannot estimate a person's depth of word knowledge in a precise and practical way. Lexical knowledge is multidimensional and so complex that research and testing methods applied so far have barely scratched the surface of the subject. Many on-line studies of the mental lexicon have concentrated on decontextualized words, and theories of bilingual lexical processing have not been accompanied by longitudinal studies illustrating in a systematic way how lexis is acquired. More needs to be done in terms of research instrument development to reflect the cumulative nature of vocabulary acquisition, a process which is different at different stages of development.

References

Aitchison, J. 1987. *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford: Basil Blackwell.

Aitchison, J. 1992. *Introducing Language and Mind*. Harmondsworth: Penguin.

Aitchison, J. 1994. *Words in the Mind: An Introduction to the Mental Lexicon*. 2nd ed. Oxford: Blackwell.

Arabski, J. 1996a. *Przyswajanie języka obcego a pamięć werbalna*. Katowice: „Śląsk".

Arabski, J. 1996b. "Stages in the acquisition of foreign lexis". In J. Arabski (ed.), *Foreign Language Acquisition Studies*. Katowice: Wydawnictwo Uniwersytetu Śląskiego, pp. 77–83.

Arabski, J. 1997. *Errors as Indications of the Development of Interlanguage*. Katowice: Uniwersytet Śląski.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Clark, E.V. 1993. *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.

Crystal, D. 2008. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.

Ellis, N. 1997. "Vocabulary acquisition: word structure, collocation, word-class, and meaning". In N. Schmitt and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Forster, K. 1976. "Accessing the mental lexicon". In R. Wales and E. Walker (eds.), *New Approaches to Language Mechanisms*. Amsterdam: North-Holland, pp. 257–287.

GABRYŚ-BARKER, D. 2005. *Aspects of Multilingual Storage, Processing and Retrieval.* Katowice: Wydawnictwo Uniwersytetu Śląskiego.

HULSTIJN, J., and M. MARCHENA 1989. "Avoidance: grammatical or semantic causes?" *Studies in second language acquisition* 11: 241–245.

KELLERMAN, E. 1977. "Towards a characterisation of the strategy of transfer in second language learning". *Interlanguage Studies Bulletin* 2: 58–145.

KELLERMAN, E. 1978. "Giving learners a break: native language intuitions as a source of predictions about transferability". *Working Papers on Bilingualism* 15: 59–92, 309–315.

KELLERMAN, E. 1985. "If at first you do succeed …". In S.M. GASS and C.G. MADDEN (eds.), *Input in Second Language Acquisition*. Rowley, Massachusetts: Newbury House, pp. 345–353.

KERSTEN, S. 2010. *The Mental Lexicon and Vocabulary Learning: Implications for the Foreign Language Classroom*. Tübingen: Narr Verlag.

KROLL, J.F. 1993. "Accessing conceptual representations for words in a second language". In R. SCHREUDER and B. WELTENS (eds.), *The Bilingual Lexicon*. Amsterdam: John Benjamins, pp. 53–81.

LAUFER, B. 1997. "What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words". In N. SCHMITT and M. MCCARTHY (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge Univesity Press, pp. 140–155.

LAUFER, B. 1998. "The development of passive and active vocabulary: same or different?" *Applied Linguistics* 19: 255–271.

LEVELT, W. 1989. *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

LEWIS, M. 1993. *The Lexical Approach*. Hove: Language Teaching Publications.

LEWIS, M. 1997. *Implementing the Lexical Approach*. Hove: Language Teaching Publications.

Longman Dictionary of Contemporary English (with CD-ROM). 2006. Harlow, Essex: Longman.

Longman Dictionary of Contemporary English. 2009. Pearson Longman.

Longman słownik współczesny. 2009. Harlow, Essex: Pearson.

MARSLEN-WILSON, W., and L. TYLER 1980. "The temporal structure of spoken language understanding". *Cognition* 8: 1–71.

MCARTHUR, T. 1981. *Longman Lexicon of Contemporary English*. Harlow: Longman.

MEARA, P. 1978. "Learners word associations in French". *Interlanguage Studies Bulletin* 3 (2): 192–211.

MEARA, P. 1982. "Word associations in a foreign language: A report of the Birbeck Vocabulary Project". *Nottingham Linguistic Circular* 11 (2): 29–38.

Meara, P. 1996. "The dimensions of lexical competence". In G. Brown, K. Malmkjaer and J. Williams (eds.), *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press, pp. 33–53.

Meara, P. 1997. "Towards a new approach to modelling vocabulary acquisition". In N. Schmitt and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 109–121.

Moon, R. 1997. "Vocabulary connections: multi-word items in English". In N. Schmitt and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 40–63.

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, P., and Waring, R. 1997. "Vocabulary size, text coverage and word lists". In N. Schmitt and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 6–19.

Oxford Wordpower. 2009. Oxford: Oxford University Press.

Oxford Advanced Learner's Dictionary. 2009. Oxford: Oxford University Press.

Paradis, M. 2004. *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Pawley, A., and F. H. Syder 1983. "Two puzzles for linguistic theory: nativelike selection and nativelike fluency". In J.C. Richards and W. Schmidt (eds.), *Language and Communication*. London: Longman, pp. 191–226.

Randall, M 2007. *Memory, Psychology and Second Language Learning*. Amsterdam: John Benjamins.

Read, J. 1997. "Vocabulary and testing". In N. Schmitt and M. McCarthy (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, pp. 303–320.

Richards, J.C. 1976. "The role of vocabulary teaching". *TESOL Quarterly* 10 (1): 77–89.

Roget's Thesaurus of English Words and Phrases. 1962. Harmondsworth, Middlesex: Penguin Books.

Rusiecki, J. 2002. "Friends true and false: or a contrastive approach to the lexicon". In J. Arabski (ed.), *Time for Words: Studies in Foreign Language Vocabulary Acquisition*. Frankfurt am Main: Peter Lang, pp. 71–81.

Schmitt, N., and M. McCarthy (eds.) 1997. *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press.

Singleton, D. 1999. *Exploring Second Language Mental Lexicon.* Cambridge: Cambridge Unviersity Press.

SINGLETON, D. 2000. *Language and the Lexicon: An Introduction*. London: Arnold.

SINGLETON, D. 2003. "Perspectives on the multilingual lexicon: a critical synthesis". In J. CENOZ, B. HUFESEIN and U. JESSNER (eds.), *The Multilingual Mental Lexicon*. Dordrecht/Boston/London: Kluwer Academic Publishers, pp. 167–176

SÖDERMAN, T. 1993. "Word associations of foreign language learners and native speakers. The phenomenon of a shift in response type and its relevance for lexical development". In H. RINGBOM (ed.), *Near-native Proficiency in English*. Abo: Abo Akademi, pp. 91–182.

URE, J. 1971. "Lexical density and register differentiation". In G. PERREN and J.L.M. TRIM (eds.), *Applications of Linguistics*. London: Cambridge University Press, pp. 43–452.

VERSPOOR, M., and W. LOWIE 2003. "Making sense of polysemous words". *Language Learning* 53: 547–586.

WESCHE, M., and T.S. PARIBAKHT 1996. "Assessing second language vocabulary knowledge: depth versus breadth". *Canadian Modern Language Review* 53: 13–40.

Wielki słownik angielsko-polski PWN-Oxford. 2002. Warszawa: Wydawnictwo Naukowe PWN.

M. KRZYSZTOF SZYMCZAK

ROZWÓJ LEKSYKALNY W PRZYSWAJANIU I UCZENIU SIĘ SŁOWNICTWA

Streszczenie

Rozdział poświęcony jest zagadnieniom związanym z przyswajaniem słownictwa obcojęzycznego. Omawia różne znaczenia takich pojęć, jak słownik (ang. *lexicon*), słownik wewnętrzny (ang. *mental lexicon*) i słowo (ang. *word*). Analizuje odpowiedź na pytanie „Co to znaczy znać dane słowo?" w odniesieniu do takich aspektów wiedzy leksykalnej, jak wymowa, pisownia, świadomość budowy wyrazu, łączenie formy słowa z jego znaczeniem, skojarzenia i związki między wyrazami, funkcje gramatyczne, kolokacje, ograniczenia co do użycia słowa. Zagadnienie słownika wewnętrznego przedstawione jest w kontekście wybranych modeli przetwarzania słownictwa. Podniesiono również kwestię integracji i rozdziału wewnętrznych słowników dwujęzycznych. Omówiono etapy przyswajania słownictwa języka pierwszego oraz zagadnienia dotyczące przyswajania słownictwa języka drugiego w nawiązaniu do wspomnianych wcześniej aspektów wiedzy leksykalnej.

M. Krzysztof Szymczak

LEXIKALISCHE ENTWICKLUNG BEI ERWERBUNG UND LERNEN DES FREMDSPRACHIGEN WORTSCHATZES

Zusammenfassung

Das Kapitel ist den mit dem Erwerb des fremden Wortschatzes verbundenen Fragen gewidmet. Der Verfasser bespricht verschiedene Bedeutung von solchen Begriffen, wie: Wörterbuch (eng.: *lexicon*), mentales Wörterbuch (eng.: *mental lexicon*) und Wort (eng.: *word*). Er versucht die Frage zu beantworten: „Was heißt, dass man ein Wort kennt?“ in Bezug auf solche Aspekte des lexikalen Wissens, wie: Aussprache, Schreibweise, Wortstrukturbewusstsein, Verbindung der Form des Wortes mit dessen Bedeutung, Assoziationen und Wortverbindungen, grammatische Funktionen, Kollokationen, Wortgebrauchsbeschränkungen. Das mentale Wörterbuch wird anhand ausgewählter Modelle der Wortschatzverarbeitung geschildert. Der Verfasser befasst sich auch mit dem Problem der Integration und Trennung von zweisprachigen mentalen Wörterbüchern. Er bespricht die einzelnen Stufen des Erwerbs von dem Wortschatz der ersten Sprache und die den Erwerb des Wortschatzes von der zweiten Sprache betreffenden Fragen.